

In search of an environmental Kuznets curve in sulphur dioxide concentrations: a Bayesian model averaging approach

JEFFREY BEGUN

*University of Washington, Box 353330, Seattle, WA 98195-3330, USA.
Email: jbegun@u.washington.edu*

THEO S. EICHER

*University of Washington, Box 353330, Seattle WA 98195-3330, USA.
Email: te@u.washington.edu*

ABSTRACT. The exact specification and motivation for an environmental Kuznets curve (EKC) is the subject of a vast literature in environmental economics. A remarkably diverse set of econometric approaches and candidate regressors have been proposed, which highlights the degree of model uncertainty surrounding the relationship between environmental quality and pollution. We introduce Bayesian model averaging (BMA) to the EKC analysis to examine: (a) whether a sulphur dioxide EKC exists, and if so (b) which income/pollution specification is supported by the data. BMA addresses model uncertainty as part of the empirical strategy by incorporating the uncertainty about the validity of competing theories into the posterior distribution. We find only weak support for an EKC, which disappears altogether when we address issues relating to the extreme oversampling of two industrialized countries in the sample. In contrast, our results highlight the relative importance of political economy and site-specific variables (specifically executive constraints and precipitation variation) in explaining pollution outcomes. Trade is shown to play an important indirect role, as it moderates the influence of the composition effect on pollution.

1. Introduction

A vast empirical literature has sought to establish a robust relationship between economic development and environmental quality. Grossman and Krueger (1995) and Selden and Song (1994) documented an inverted U-shaped curve between income and pollution that is similar to Kuznets's (1955) inverted U-shaped relationship between income and inequality. In subsequent research, a large number of authors failed to confirm an 'environmental Kuznets curve' (EKC), either in the original Grossman and Krueger dataset, or in updated and expanded pollution datasets (e.g., Harbaugh *et al.*, 2002, or Deacon and Norman, 2006). The conflicting empirical results have given rise to intense efforts to further explore the income/pollution relationship either by introducing formal models (see,

We thank Werner Antweiler and Bill Harbaugh for sharing their data, and Chris Papageorgiou and three referees for extensive and helpful comments.

e.g., Antweiler *et al.*, 2001), or by adding further control variables to reduced-form regressions (see Dasgupta *et al.*, 2002 for a survey).

The EKC is thus a case study of extreme model uncertainty where the true model is unknown and several competing approaches exist that hypothesize about the exact relationship between environmental quality and income. In light of such model uncertainty, inference procedures based on a single regression model overstate the precision of coefficient estimates. Precision is overestimated because the uncertainty surrounding the validity of a theory has not been taken into account (Raftery, 1995). The problem is particularly prevalent in the EKC literature since a number of well-founded approaches exist and researchers face an abundance of possible candidate regressors.

The Bayesian solution to model uncertainty is to base inferences on all competing models, each weighted by the posterior probability that the model is indeed the true model. The procedure delivers a posterior distribution for each candidate regressor, whose mean is a weighted estimate derived from all relevant models. In environmental economics, prominent examples of BMA applications include the modeling of population determinants for deer (Farnsworth *et al.*, 2006), fish (Fernandez *et al.*, 2002) as well as pollution mortality (Koop and Tole, 2006). To our knowledge, we are the first to apply Bayesian model averaging to resolve the model uncertainty surrounding the EKC relationship.

Our strategy is to group EKC approaches into two categories. First we examine reduced-form approaches to the EKC, where many possible determinants of pollution are introduced. This branch of the literature is vast, but suffers from the criticism that the direct and indirect effects of variables cannot be disentangled. The approach therefore cannot identify intervening factors that lead to an apparent relationship between income and pollution. As an alternative, we examine specific theories that have been proposed as the underlying determinants of an EKC, and scrutinize whether the data support theory-based candidate regressors.¹ In this case, we have a clearly predetermined set of regressors that are expected to affect pollution concentrations.

Before we summarize our results, it is important to note that the updated SO₂ data that have been extended and cleaned of previous errors no longer exhibit the EKC relationship that Grossman and Krueger (1995) discovered (see, e.g., Harbaugh *et al.*, 2002). Our results below can therefore be seen as an effort to find robust evidence for an EKC in this dataset by eliminating possible omitted variable bias. We find only limited evidence for an income/pollution relationship once we account for model uncertainty. Instead, robustly related regressors in both reduced-form and theory-based approaches are those relating to political economy, site-specific effects, and trade (the individual proxies for each category are motivated in sections 3 and 4). Societies that are more open in terms of political participation are shown to exhibit significantly lower air pollution. The theory-based approach highlights the power of both direct and indirect effects (where

¹ Theory-based in this context refers to regressors that have been proposed in the past to test these specific theories. Some regressors have been motivated by several alternative theories.

indirect refers to interactions where one variable moderates the effect of another variable). Following Antweiler *et al.* (2001), we show that the interaction between trade and capital intensity is also of crucial importance for explaining the evolution of SO₂ concentrations across countries and time.

The number of regressors that are robustly related to pollution in the BMA approach, as well as in the best model identified by BMA, is only a fraction of the 17 possible candidate regressors motivated by reduced-form approaches. Compared to our selection of theory-based specifications, BMA finds as few as a third of the 18 regressors suggested by the most comprehensive theory-based specification. Nevertheless, the best model suggested by BMA has an adjusted R^2 three times greater than the preferred theory-based specification of Antweiler *et al.* (2001).² This provides evidence that such a complex theory may not be necessary and alternative theories, such as the Green Solow model (see Brock and Taylor, 2005), should not be discarded simply because they rely on only a fraction of the regressors.

2. Searching for an EKC in SO₂ concentrations

2.1 Data considerations

One prominent EKC relationship in the literature relates air quality to economic development.³ In this paper we focus on sulphur dioxide (SO₂) concentrations obtained from the Global Environmental Monitoring System (GEMS). The data are updated, error-corrected, and maintained by the EPA in its Aerometric Information Retrieval System (AIRS).⁴ The GEMS/AIRS data are perhaps the most widely used dataset to investigate the EKC, with reported SO₂ concentrations from stations in up to 44 countries from 1971 to 2006.⁵

² The comparison of different BMA specifications in this paper may involve not only different regressors, but also different samples sizes. Adjusted R-squared compensates for both by expressing the explained variation after accounting for differences in sample sizes and the number of regressors. More complex measures for model comparison are available that balance the opposing pressures of goodness of fit and complexity (see Burnham and Anderson (2002) for an exhaustive discussion). One might be suspicious that sample size was driving our results if models with larger samples sizes performed better. The opposite is the case here.

³ Alternative measures have been used. Evidence for an EKC has been found, for example, for water quality (Grossman and Krueger, 1995), deforestation (Cropper and Griffith, 1994; Panayotou, 1995), and water withdrawal for agriculture (Rock, 1998; Goklany, 2002). Some researchers have found an EKC for carbon dioxide (Roberts and Grimes, 1997) though others have found that CO₂ increases monotonically with income (Shafik and Bandyopadhyay, 1992).

⁴ Our raw GEMS/AIRS data are identical to that of Antweiler *et al.* (2001), who kindly shared their data that include median concentrations.

⁵ See, for example, Grossman and Krueger (1995), Panayotou (1997), Torras and Boyce (1998), Barrett and Graddy (2000), Harbaugh *et al.* (2002), and Deacon and Norman (2006). De Groot *et al.* (2004) use concentrations/GDP.

Our income measure is real GDP per capita in constant 1996 dollars from the Penn World Tables 6.1 (Heston *et al.*, 2002). In our estimation of the Antweiler *et al.* (2001) approach we use their income measure (GNP). There are several reasons to use concentrations data, although emissions data are also widely available. First, ground-level sulphur dioxide concentrations are the relevant criterion for the direct environmental/health impact. Second, cross-country emissions data are generated by emissions models and not based on actual SO₂ measurements. Cross-country emissions data are generated using strict input–output coefficients based on energy and manufacturing models.⁶ In that sense these data perfectly track the production characteristics of an economy, but not necessarily the actual factors that affect environmental quality. Third, the majority of SO₂ EKC papers use concentrations and we seek to be comparable with our results.

The SO₂ concentrations data are, however, highly unbalanced in two dimensions: location and time. Few countries report data over the entire time period, and many countries report pollution concentrations for less than a decade. Often several years of data are missing between observations not only on the station level, but also on the city and country levels. Even in countries with extensive locational coverage, such as the United States, the time series for each monitoring station is highly unbalanced. When heavily oversampled countries have lower average pollution, or have added new monitoring stations with lower pollution over time, it is important to examine the robustness of the results at different levels of data aggregation.⁷ Juxtaposing different levels of aggregation as well as reduced-form and structural results also provides a unique opportunity to examine the robustness of regressors across all specifications.

A significant number of papers in the literature have documented an EKC (or its absence) without explicitly discussing the fact that the dataset is so extremely unbalanced. The data are unbalanced in terms of location since a few countries are represented with a large number of reporting stations, while many other nations are featured only once. A full 38 per cent of the original 2,555 station-level observations originate in the US and Canada. The imbalance is exacerbated early and late in the sample as the US supplies 69 per cent of the data before 1974 and after 1993. Therefore, we restrict our analysis to 1974–1993, which reduces the dataset by 219 observations (almost exclusively from the US). Figure 1 provides a breakdown of the 2,168 observations by country of origin.

The construction of the appropriate station-level covariates is also problematic. None of the covariates suggested by the literature actually speaks to specific characteristics at the station level. At best one can

⁶ The data generate ‘estimates’ of SO₂ emissions based on national statistics and the use of sulphur bearing fuels and metals, weighted by their sulphur content.

⁷ The alternative approach to oversampling would be to use data weighting schemes that rebalance the data so that, for example, data from heavily oversampled countries receive a lower weight. Such schemes require knowledge of the true sampling distribution, however, and knowledge of the determinants and structure of the selection probabilities for each station (e.g., why certain country’s states were so much more likely to be sampled).

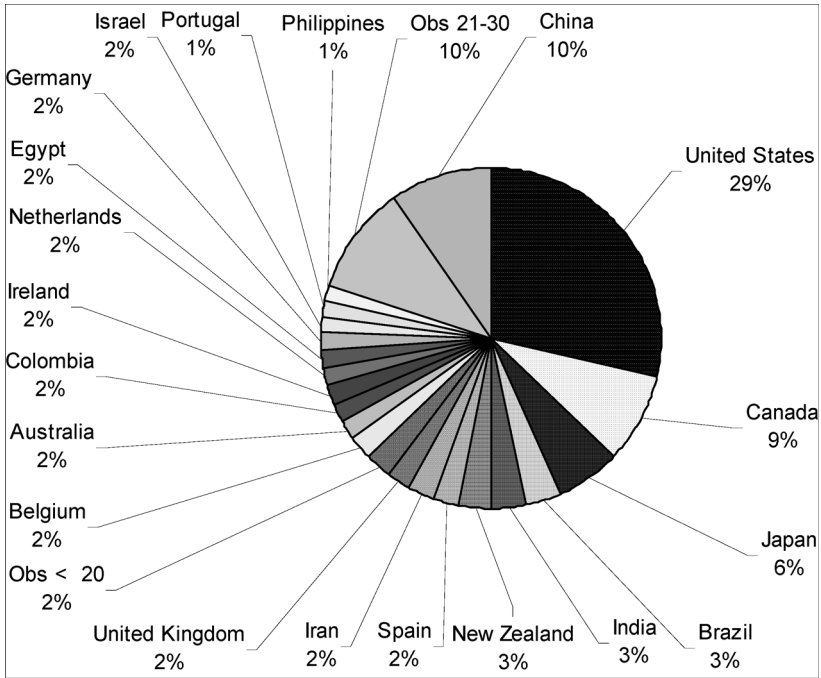


Figure 1. Distribution of SO₂/observations in GEMS/AIRS data 1974–1993

Notes: Obs < 20 refers to a number of countries that feature less than 20 observations (Greece, Italy, Thailand, Venezuela, Malaysia, Indonesia, South Korea, Peru, Pakistan). Obs 21–30 refers to countries with between 21 to 30 observations (Argentina, Finland, Sweden, Chile, France, Denmark).

Source: US–EPA maintained GEMS/AIRS dataset <http://www.epa.gov/airs/aexec.html>

correlate city-level characteristics (such as temperature and precipitation variation) with observed concentrations. However, even the city-level data are still highly unbalanced, and it is unclear whether results are driven by information in the data or by oversampling and missing information across time. In that sense, aggregating to the country level is our preferred approach; it suffers, however, from the disadvantage that few of the controls actually speak to the characteristics of the area in which the concentrations are measured. For these reasons we test for an EKC at the station, city, and country levels.

2.2 The EKC in the raw income pollution data

The first surprise for researchers using the newest version of GEMS, which has been purged of errors and extended to include updated data, is that it no longer provides evidence for the fundamental EKC relationship. Figure 2 plots the raw data for every station in every year that an observation is recorded. In addition, the figure traces the predicted values from the most fundamental regression that includes only log median SO₂ concentrations

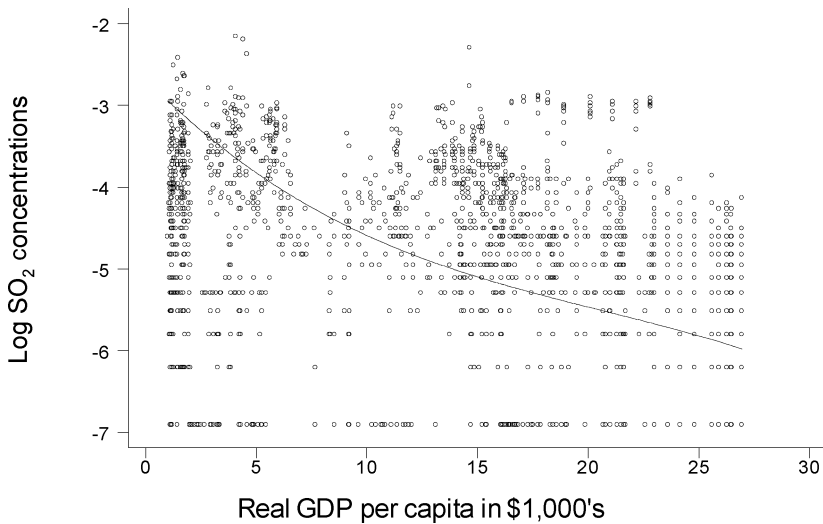


Figure 2. Relationship between median SO_2 concentrations and income (by measuring station 1974–1993)

Note: Fitted values are for the fixed-effects regression $\text{Log } SO_{2it} = \alpha_i + \beta \text{GDP}_{it} + \delta \text{GDP}_{it}^2 + \gamma \text{GDP}_{it}^3 + \varepsilon_{it}$.

Source: US–EPA maintained GEMS/AIRS dataset <http://www.epa.gov/airs/aexec.html>

as the dependent variable and real GDP per capita as a third-order polynomial.⁸ In Grossman and Krueger (1995), a similar plot using earlier data from the same source was prominently inverted-U shaped.

Instead of an EKC, the updated GEMS data in figure 2 show a simple relationship between development and environmental quality that has SO_2 concentrations gradually declining with income. The lack of an EKC in the raw SO_2 data has previously been noticed (on the country level) by astute researchers who suggested that the global data mask country-level phenomena. Deacon and Norman (2006) provide evidence for 23 countries that the country-level experience may in fact look very different from the original global station-level data. Since technology, factor abundance and the political response to interest groups are also national concepts, we aggregate the data in search of an EKC at the individual country level.

⁸ We employ fixed-effects regressions throughout. It has been argued that the random-effects EKC cannot be estimated consistently (Mundlak, 1978; Hsiao, 1986; Stern, 2004). Since the very premise of the EKC is that specific local, regional, or national characteristics are crucial, the random-effects approach suffers from inconsistency due to omitted variables. In addition, we have no desire to imply that a possible EKC in our data holds beyond the countries in this sample. Hence, we take the view that represented countries are not simply random draws from a larger EKC country/station population. An additional advantage of the fixed-effects approach is that it controls for many time-invariant, site-specific, and country-specific factors.

Plotting country-level SO₂ concentrations over time confirms Deacon and Norman's (2006) result that most countries' SO₂ concentrations do not follow an EKC path.⁹

The lack of an EKC at the station or individual country level might also be an artifact of the extremely unbalanced time dimensions of the dataset. To balance the sample intertemporally, we follow Selden and Song (1994) and take five-year averages.¹⁰ In the averaged dataset, the US prominence is reduced to 24 per cent of the observations at the station level.¹¹ Therefore, averaging helps address our oversampling concerns, and in the country-level data the entire locational imbalance that leads to oversampling concerns is eliminated. Averaging across time and aggregating by country does not resolve the mystery of the missing EKC in the raw data, however. Plotting station-level data and predicted values obtained by the same method as in figure 2, the country-level data in figure 3 maintain the negative relationship between pollution and income.

3. Model uncertainty in the income/environment relationship

Two simple explanations can address the absence of an EKC in the raw data presented in figures 2 and 3. Either the relationship does not exist, or the model is misspecified. By neglecting to include crucial covariates, the misspecification due to omitted variable bias may overwhelm the power of the GDP regressors. Perhaps in an effort to explore the latter line of reasoning, a number of papers in the literature feature a remarkably diverse range of different model specifications to uncover evidence in favor of an EKC.

Below, we first focus on the most prominent reduced-form approaches that commonly include variables to sharpen the EKC model specification such as international trade, capital intensity, precipitation variation, temperature, population density, investment, education, and institutions. These diverse approaches represent the level of model uncertainty that surrounds the EKC relationship. Standard robustness analysis would juxtapose various models and select on the basis of P-value. As Miller (1984, 1990) points out, the difficulty is that a P-value based on a model selected from a larger set of possibilities no longer carries the same interpretation as when only two models are considered (the null and the alternative). Also, several models may seem reasonable given the data but lead to different conclusions. This can happen especially in cases when the dataset

⁹ The plot is included in the working paper version of this paper (Begun and Eicher, 2007).

¹⁰ Averaging data over five-year intervals is also common in the economic growth literature and allows us to address the error associated with business cycle fluctuations that are inherent in income data (see Barro, 1990).

¹¹ Countries that lack SO₂ data for at least two five-year periods are excluded. These countries (and their number of observations) are Austria (2), Kenya (4), Switzerland (2), Ghana (3), and the Czech Republic (21). Poland (86 observations), and Iraq (9 observations) are excluded from the reduced-form analysis since we lack PWT 6.1 GDP data for these countries. Hong Kong (40 observations) is excluded because it lacks Polity IV data, and Yugoslavia (63 observations) was dropped by Antweiler *et al.* (2001) because it lacks human capital data.

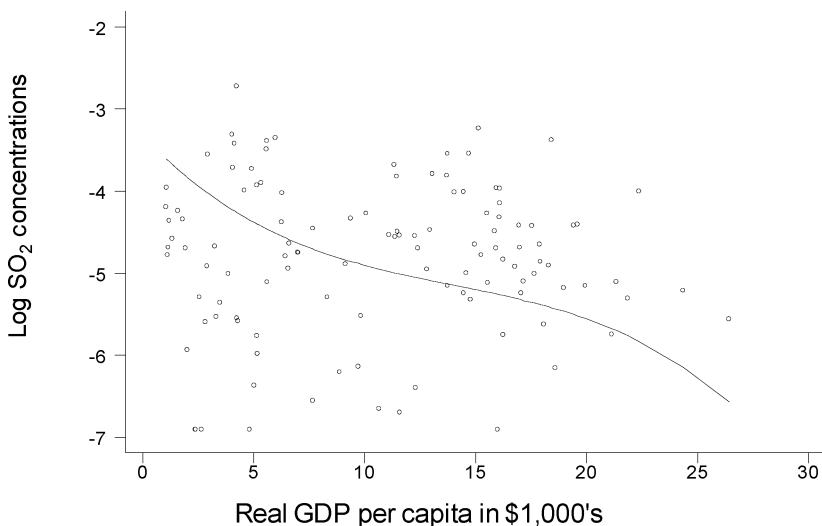


Figure 3. Relationship between median SO₂ concentrations and income (five-year averages 1974–1993)

Note: Five-year averages of log median SO₂ concentrations, aggregated from the station to the country level.

Source: US–EPA maintained GEMS/AIRS dataset <http://www.epa.gov/airs/aexec.html>

is large (for striking examples see Kass and Raftery, 1995; Raftery, 1996). The Bayesian approach to model selection and accounting for model uncertainty overcomes these difficulties. The next subsection provides a brief overview of BMA and identifies how the procedure addresses EKC model uncertainty.

3.1 Addressing model uncertainty in the income/environment relationship

When inferences are based on one model alone, the ambiguity involved in model selection dilutes information about effect sizes and predictions since ‘part of the evidence is spent to specify the model’ (Leamer, 1978: 91). Model averaging was first operationalized by Leamer (1983) in so-called ‘extreme bound analysis’ (EBA). EBA has two limitations. First, in the absence of an efficient search, EBA arbitrarily restricts the set of candidate regressors (and hence the model space).¹² EBA it is not anchored in foundations of statistical theory and in practical applications it has been shown to be biased towards selecting too few ‘effective’ regressors (see Sala-i-Martin, 1997; Sala-i-Martin *et al.*, 2004).

BMA inference is based on an unrestricted search of the model space spanned by all candidate regressors. BMA also requires that each model is weighed according to its quality. This quality weight is given by the

¹² Depending on the search algorithm, BMA might also restrict the model size (see Eicher *et al.*, 2007a). However, the Raftery *bicreg* algorithm is based on the *leaps and bounds algorithm*, which does not restrict the model space in our particular application.

posterior model probability, which is interpreted as the probability that any given model is the true model. Extreme bound analysis weighs models equally and thus attributes equal power of inference to exceptionally weak or strong models. Sala-i-Martin (1997) does introduce an *ad hoc* weighting scheme to BMA; his results highlight the sensitivity of EBA to the weights, and therefore the need to derive such weights using actual statistical theory. Hjort and Claeskens (2003) point out that for good reasons BMA 'dominates the literature on accounting for model uncertainty in statistical inference'. Raftery and Zheng (2003) summarize the main theoretical results proving that BMA: (a) minimizes the total error rate (the sum of type I and type II error probabilities); (b) produces point estimates and predictions that minimize mean squared error (MSE); and (c) yields predictive distributions that have optimal predictive performance relative to other approaches. The authors also outline the differences between Bayesian model averaging and frequentist model averaging, as well as the conceptual problems involved in frequentist model averaging.

It is therefore not surprising that averaging over all models can be analytically proven to provide better average predictive performance than any given regression, any single selected model (using selection procedures such as stepwise regression), or any subset of models (Madigan and Raftery, 1994). Eicher *et al.* (2007a) provide concrete examples of this phenomenon using growth and simulated data and show not only that BMA attains the theory-predicted superior inference, but also that the quality of models discovered by alternative methods, such as the 'general-to-specific' (GETS) procedure (suggested by Hendry and Krolzig, 2001) is far inferior to BMA's.

The basic model averaging idea originated with Jeffreys (1961) and Leamer (1978), whose insights were developed and operationalized by Draper (1995) and Raftery (1995). BMA was first introduced to economics by Fernandez *et al.* (2001), with an application to economic growth. Here we restrict ourselves to sketching the basic BMA structure before we discuss the results (for an extensive discussion of BMA see Hoeting *et al.*, 1999).

The basic variable selection setup can be concisely summarized as follows. Given a dependent variable, Y (SO_2 concentrations), a number of observations, n , and a set of candidate regressors, X_1, X_2, \dots, X_k , the variable selection problem is to find the 'best' model

$$Y = \alpha + \sum_{j=1}^p \beta_j X_j + \varepsilon, \quad (1)$$

where X_1, X_2, \dots, X_p is a subset of X_1, X_2, \dots, X_k , and β_j is a vector of regression coefficients to be estimated. Let $M = \{M_1, \dots, M_k\}$ denote the set of all models considered, and let $\theta_k \sim (\beta_k, \sigma^2)$ be a vector of parameters in M_k . The likelihood function of model M_k , $pr(D|\theta_k, M_k)$, given the data, D , then summarizes all information about θ_k that is provided by the data.

For any likelihood function consisting of two or more parameters, we can define the *integrated likelihood* as the probability of the data given model M_k . The integrated likelihood of model M_k , $pr(D|M_k)$, is the likelihood function

times the prior density, $pr(\theta_k|M_k)$, integrated over the parameters¹³

$$pr(D|M_k) = \int pr(D|\theta_k, M_k) pr(\theta_k|M_k) d\theta_k. \quad (2)$$

The integrated likelihood is the crucial ingredient in deriving the appropriate model weight used in the model averaging process. Given the prior probability that M_k is the true model, $pr(M_k)$, the posterior probability of a model, $pr(M_k|D)$, is defined as the model's share in the total posterior mass

$$pr(M_k|D) = \frac{pr(D|M_k) pr(M_k)}{\sum_{l=1}^K pr(D|M_l) pr(M_l)}. \quad (3)$$

Equation (3) thus represents the individual model's weight in the averaging process. Posterior model probabilities are also the weights used to establish posterior means and variances, which have been derived by Raftery (1993)

$$E[\beta|D] = \sum_{k=0}^K \hat{\beta}_k pr(M_k|D), \quad (4)$$

$$Var[\beta|D] = \sum_{k=0}^K (Var[\beta|D, M_k] + \hat{\beta}_k^2) pr(M_k|D) - E[\beta|D]^2, \quad (5)$$

where $\hat{\beta}_k$ is the OLS estimate for M_k , $\hat{\beta}_k = E(\beta|D, M_k)$. Hoeting (1994) derives the full expression for the definitive posterior distribution. Hence, the posterior means and variances are simply the first and second moments of each individual model, weighted by the model's 'quality', as given by its posterior probability. BMA thus incorporates model uncertainty into the posterior distribution such that the variance of the weighted model average is greater than the variance for any single model as long as there is disagreement across models. Intuitively, the different models are used to describe different parts of the data, rather than to pretend that a single model can describe all the data. An individual model does not account for the uncertainty about the model actually being the true model, and hence parameter estimates' variances overstate the confidence in the estimate.¹⁴

¹³ The software used to implement BMA is available at <http://www.research.att.com/~volinsky/software/bicreg>. It utilizes the *unit information prior*, which is so diffuse and uncontroversial that it can be derived from frequentist statistics. It is a multivariate normal prior whose mean is centered at the maximum likelihood estimate and whose variance equals the expected information matrix for one observation (Kass and Wasserman, 1995).

¹⁴ An additional dimension of model uncertainty relates to the uncertainty regarding the correct econometric approach. BMA cannot resolve this uncertainty. For example, List and Gallett (1999) and Millimet *et al.* (2003) explore non-parametric modeling strategies using emissions rather than concentrations. Martínez-Zarzoso and Bengochea-Morancho (2004) use a mean group approach that consists of estimating separate regressions for each country and calculating averages of the country-specific coefficients in the context of CO₂ pollution. Dijkgraaf and Vollebergh (2005) also reject an EKC in CO₂ and find homogeneity among countries to be a problematic assumption. Alternative modeling strategies usually compare their results to parametric approaches. Unfortunately, a full-fledged comparison

In addition to the posterior means and standard deviations, BMA provides the *posterior inclusion probability* of a candidate regressor, $pr(\beta_k \neq 0|D)$, by summing the posterior model probabilities across those models that include the regressor. Posterior inclusion probabilities provide a probability statement regarding the importance of a regressor that directly addresses researchers' prime concern: what is the probability that the coefficient has a non-zero effect on the dependent variable.¹⁵

4. Motivating EKC candidate regressors

Before we can employ BMA, each candidate regressor must be motivated to justify its inclusion alongside GDP measures, since each regressor can only be included if it corresponds to a well-established theory or line of research. As mentioned in the introduction, a regressor may be motivated by several theories. Numerous covariates have been introduced in the past to explain sulphur dioxide concentrations in reduced-form specifications. These regressors can be grouped into five different categories: (1) site-specific controls, (2) political economy proxies, (3) production structure, (4) trade measures, and (5) technology proxies. Note that references to our selection of theory-based approaches are limited in this paper to single equation specifications. We ruled out theory-based approaches based on energy demand decomposition that underlie the emissions literature (see Ang, 1999, and Hoekstra and Van den Bergh, 2002 for reasons explained in our discussion on emissions data above). However, the distinctions between separate effects suggested by theory (e.g., scale, technology, trade) can also be explored using a system of equations that explores the dynamics interaction and transition of the variables (see, e.g., Stern, 2005; Constantini and Martini, 2007). The limitation of BMA is the single equation approach. Reduced-form approaches constitute the vast majority of EKC papers (Stern, 2004); they cannot, however, identify the true effect of regressors, be it direct or indirect. This requires fully specified models.

Since concentrations are reported at the station level, a compelling argument can be made that any analysis of the income–pollution relationship must include regressors that control for site-specific factors (e.g., temperature and precipitation variation). Such regional differences

of the parametric BMA results vs. alternative modeling strategies on the basis of predictive performance along the lines of Eicher *et al.* (2007a) is beyond the scope of this paper. Alternatives might be to explore a parametric approach together with parameter heterogeneity. This approach has been examined in the growth context by Eicher *et al.* (2007b), where theory identifies clear groups of heterogeneous countries (e.g. Africa or Asia). We are not aware that the environmental quality literature features such an established set of country clusters to test for slope heterogeneity.

¹⁵ The posterior inclusion probability thus also carries an important interpretation that is more informative than standard P-values. P-values are often interpreted as indicating that a coefficient has a chance of being equal to zero, or that the null hypothesis has a chance of being true. However, P-values only indicate something much more obscure: if the null hypothesis, $\theta = 0$, were true, then the probability of collecting data as extreme as or more extreme than what is observed is equal to the P-value.

affect nature's ability to cleanse SO₂ from the atmosphere. While variables such as temperature and precipitation variation are unlikely to be correlated with our economic variables, their inclusion is standard in the literature and meant to improve the accuracy of the estimates. Our site-specific controls for temperature and rainfall are obtained from Antweiler *et al.* (2001).¹⁶

Aside from station characteristics, we must also control for effects that are common-to-world but nevertheless time varying. Such components reflect secular changes in global awareness of environmental problems, innovations, diffusion of technology and the evolution of world prices. We follow the standard practice in the literature and assume that these common components are captured by a linear time trend. In addition, we add a dummy for nations that signed the 1985 Helsinki Protocol, which aimed to reduce sulphur dioxide emissions by at least 30 per cent.

Income alone does not create direct pressure to improve environmental outcomes, but the democratic fabric of a society that allows political participation and threatens consequences for polluting dictators has been found to be an important determinant. The past literature introduced variables that indicate when more open and democratic societies have different attitudes towards the environment. The conjecture is that for a given level of income, more open societies experience less pollution.¹⁷ Torras and Boyce (1998) posit that richer individuals gain 'power' to demand better overall environmental quality. Likewise, Barrett and Graddy (2000) propose that wealthier citizens demand an increase in the non-material aspects of their standard of living. The degree to which policy responds to such desires is closely linked to the ability of individuals to assemble, organize and voice their concerns. In the same vein, Panatayotou (1997) provides evidence that strong property rights 'flatten' the EKC by generating less pollution for any given income level. Some authors employ the Freedom House indices to measure political rights (e.g., Shafik and Bandyopadhyay, 1992; Torras and Boyce, 1998; Barrett and Graddy, 2000), while others use Knack and Keefer's (1995) 'Respect/Enforcement of Contracts' (Panatayotou, 1997). Harbaugh *et al.* (2002) use an index of democratization from the Jagers and Gurr (1995) Polity III dataset. Alternatively, Leitão (2006) introduces measures of corruption.

The institutions and growth literature has since established the Polity IV 'Constraint on Executive' (Marshall and Jagers, 2003) as the best measure to capture the above-mentioned effects. Acemoglu *et al.* (2001) have shown

¹⁶ Grossman and Krueger (1993, 1995) also include site-geography variables such as proximity to oceans or deserts. Our fixed-effects regressions account for these implicitly.

¹⁷ Of course one could also conjecture that income is simply an *indicator* for deeper mechanisms that drive environmental quality. Thus, exploring all controls that have been mentioned in the past literature, *other* than income, is an important robustness analysis. Since BMA explores the entire model space, it is uniquely qualified to identify any combination of such deeper mechanisms, which could reduce and even negate the explanatory power of income – if income is only an indicator. If income remains significant, we either have not discovered the correct mechanism, or income may not be a simple indicator.

convincingly that the degree of constraint on the executive is a fundamental determinant of all political rights. We thus choose this measure as our political rights proxy. Antweiler *et al.* (2001) include a site-specific dummy for Communist regimes, which is interacted with per capita income as a political proxy. We leave it to BMA to identify whether executive constraints or the Communist regimes dummy proxies most effectively for political economy effects.

Since a key hypothesis is that political pressure builds as richer agents demand greater environmental quality, education is also seen as a major factor in the pollution/development relationship. Torras and Boyce (1998) include adult literacy rates, noting that literacy allows for greater informational access and a more even distribution of power within society. Our measure of education is years of education from Barro and Lee (2000). Years of education should be a better proxy for access to information since basic literacy implies only knowledge of rudimentary reading and writing skills. We use average years of education over the prior three years to account for the fact that it takes some time to translate educational achievement into environmental activism.¹⁸

International trade has also been associated with the EKC relationship. Arrow *et al.* (1995) and Stern *et al.* (1996) mention that an EKC might be partly due to trade and the resulting global distribution of polluting industries. The authors hypothesize that free trade allows developing countries to specialize in goods that are intensive in their relatively abundant factors: labor and natural resources. Developed countries, in turn, are likely to specialize in human capital and capital intensive goods. In contrast, Shafik and Bandyopadhyay (1992) point out that trade might exert two contrasting influences on developing countries. Following Antweiler *et al.* (2001), we use trade volume (exports plus imports) as a per cent of GDP as our measure of openness to trade. Aside from the above-mentioned trade effect, increased openness may lead to increased competition, which could cause more investment in efficient and cleaner technologies to meet the environmental standards of developed nations. More directly, investment can be motivated with reference to embodied technology, where cleaner technologies are embodied in more recent vintages of capital. To control for such an effect, we follow Harbaugh *et al.* (2002) and include not only trade, but also a measure of investment in our analysis. Alternatively, trade-induced dynamic comparative advantage has also been tied to the composition of output that is associated with different stages of development. We use the human-capital-adjusted capital intensity proxy from Antweiler *et al.* (2001) to account for such effects.

Another important covariate often included in the literature is population density (Grossman and Krueger, 1995; Panayotou, 1997; Barrett and Graddy, 2000; Antweiler *et al.*, 2001; Harbaugh *et al.*, 2002). Panayotou argues that

¹⁸ Barro and Lee (2000) is the most commonly used dataset for international comparisons. While often criticized, no better global dataset has been developed. For OECD countries de la Fuente and Doménech (2001, 2006) show that one can increase the signal-to-noise ratio by exploiting additional data sources, and by eliminating breaks in the data that seem to be a function of data collection.

population density may have an ambiguous effect since more dense areas can expect greater use of coal and non-commercial fuels, but densely populated countries may also be more concerned about lowering pollution concentrations. We follow Harbaugh and include national population density in order to have a relatively accurate time-series measure of population density for both developed and developing countries.

Second-generation EKC models include variables motivated by fully specified models that yield precise, testable EKC implications and relationships. The essential features of EKC models include determinants of scale, composition, and technique effects outlined by Panayotou (1997). Prominent theoretical precursors that have led to the state of the art, fully specified, open-economy EKC model in Antweiler *et al.* (2001) are Stokey (1998) (endogenous abatement), Bovenberg and Smulders (1995) and Aghion and Howitt (1998) (endogenous growth/technique), and Jones and Manuelli (2001) (endogenous policy). Development causes a positive scale effect since increased output per unit of capital leads to increased pollution. To account for the scale effect we follow Antweiler *et al.* (2001) and employ a measure of city-level economic intensity (national GDP per capita times city population density). It is generally held that in rapidly growing middle-income countries pollution due to the scale effect might be the dominant EKC force (Perman and Stern, 2003).

The technique effect diminishes the scale effect as technological progress permits a lowering of emissions per unit of output, which presumably would also impact SO₂ concentrations. Lagged per capita income is used to proxy for the technique effect since countries with higher incomes in the past should be able to afford better technology today (see Antweiler *et al.*, 2001). Diffusion of technology itself motivates the idea that time-related effects reduce environmental impacts in countries at all levels of development (Aghion and Howitt, 1998; Perman and Stern, 2003).¹⁹ These effects are usually proxied with a year dummy.

To isolate either the scale or technique effect, we must control for changes in the composition of output. A change in output composition can mitigate the scale effect further if the share of less pollution-intensive industries rises as income increases. This occurs when development and human capital accumulation generate shifts toward cleaner industries (services or information technology) so that the ensuing change in the composition of output reduces environmental degradation (Panayotou, 1993). A specific model was first presented by Copeland and Taylor (2003), who showed that the reliance on capital accumulation in early stages of development, as opposed to human capital accumulation in later stages, can generate an EKC. Following Antweiler *et al.* (2001), we capture the composition effect by controlling for differences in the human-capital-adjusted capital-labor ratio. In the absence of such controls, the relationship between pollution

¹⁹ In contrast to simple fixed- and random-effects regressions, where collinearity between GDP and lagged GDP variables might compromise the explanatory power of either variable, BMA averages across relevant models and thus potentially mitigates the effects of collinearity.

and income is a mixture of scale, composition, and technique effects, which is hard to interpret.

In addition to a simple income/output-induced composition effect, Antweiler *et al.* (2001) also account for a trade-induced composition effect. While the reduced-form literature takes the effect of trade as ambiguous, Antweiler *et al.* (2001) stipulate that a trade-induced composition effect depends on a country's comparative advantage, which in turn is determined by income per capita and capital abundance. Antweiler *et al.* (2001) suggest interacting the trade measure with determinants of comparative advantage (the capital-labor ratio and income per capita). Since comparative advantage is a relative concept, these variables are measured relative to their corresponding world averages. Since theory cannot identify a turning point (the endowment levels where trade causes a switch from exporting to importing pollution-intensive products), we adopt the flexible approach of Antweiler *et al.* (2001) by estimating interactions with different functional forms.

Note that theory does not necessarily imply such an elaborate structure. Brock and Taylor (2005) point out that the EKC is compatible with many different theories. The simplest of all is perhaps the 'Green Solow model' where pollution policy remains unchanged throughout the development process and where transitional dynamics alone suffice to generate an EKC. The Green Solow model exhibits no composition effects, no changes in pollution abatement, no evolution of the political process, and no international trade. BMA is a natural statistical tool to examine the support that competing theories receive from the data, and to address the model uncertainty in the literature.

5. Empirical results

Tables 1–3 report the reduced-form results, while table 4 reports the results for our candidate regressors that were motivated by our selection of theory-based approaches. The main results are robust to specifications of GDP in logs, different GDP lag structures (zero-, three-, and ten-year lags), alternative 'U-curve' specifications such as Anand and Kanbur's (1993) specifications based on inverse GDP, and specifications based on concentrations per capita. The tables report results at the station, city, and country levels.

The first columns of tables 1–3 report the posterior inclusion probability (the probability that the coefficient estimate is different from zero). $P \neq 0$ is thus a measure of confidence that a regressor enters with a non-zero coefficient into the true regression model. The posterior inclusion probability is a scale-free probability measure of the relative importance of variables; it can therefore be transparently applied to inform policy decisions, in addition to the posterior mean and standard deviation. Jeffreys (1961) and Raftery (1995) add the interpretational refinement that $P \neq 0 > 50$ per cent indicates that the data provide weak evidence that a regressor is included in the true model; $P \neq 0 > 75$ per cent implies positive evidence; $P \neq 0 > 95$ per cent provides strong evidence; and $P \neq 0 > 99$ per cent gives very strong evidence. Inclusion probabilities close to 100 per cent signal that a particular regressor is included in almost all good models so that it

Table 1. *Reduced-form BMA results (by station)*

	<i>Dependent variable: log median SO₂ concentrations, 5-year averages</i>		
	<i>P</i> ≠ 0	<i>Posterior mean (S.D.)</i>	<i>Best model mean (S.D.)</i>
Intercept	100.0	-2.5805 (3.0721)	-2.4109* (1.3777)
Trade	100.0	-2.8938 (0.4317)	-2.9285*** (0.4922)
Executive constraints	100.0	-0.1926 (0.0307)	-0.1972*** (0.0380)
Precipitation variation	99.9	44.5363 (9.1457)	45.1435*** (10.6779)
(GDP _{<i>t</i>-3}) ²	79.8	-2.8297 (1.6156)	-3.4595*** (0.9308)
(GDP _{<i>t</i>-3}) ³	79.5	0.5284 (0.3044)	0.6416*** (0.1790)
GDP _{<i>t</i>-3}	79.4	4.4384 (2.5736)	5.5128*** (1.6123)
Temperature	78.3	-0.1396 (0.0909)	-0.1799** (0.0785)
Investment	44.9	-0.0087 (0.0110)	.
(GDP) ²	21.2	-0.4925 (1.1662)	.
(GDP) ³	20.6	0.0849 (0.2090)	.
GDP	19.2	0.8001 (1.9521)	.
Helsinki	12.9	-0.0370 (0.1152)	.
Capital intensity (H*K/L)	10.4	0.0094 (0.0335)	.
(Capital intensity, H*K/L) ²	9.1	0.0004 (0.0016)	.
Education _{<i>t</i>-3}	6.1	0.0042 (0.0223)	.
Year	3.1	-0.0001 (0.0014)	.
National population density	2.4	0.0010 (0.0497)	.
Observations			623
Adjusted R ²			0.240

Notes: *P* ≠ 0 is the posterior inclusion probability that a regressor's posterior mean is different from zero. *, **, ***, indicate 90, 95, 99 per cent confidence levels.

contributes prominently to explaining the dependent variable, even in the presence of significant model uncertainty.

We find only limited support for income as a key driver of SO₂ concentrations. Only the highly unbalanced station- and city-level datasets in tables 1 and 2 report positive evidence of an EKC relationship between income and SO₂ concentrations. At the station level, lagged GDP has a much higher inclusion probability than current GDP, implying that contemporaneous economic activity is less important in determining SO₂ concentrations than the indirect effects of rising income over time. Nevertheless, fundamental variables, not income, are the crucial determinants of pollution levels. Precipitation variation and executive constraints both exhibit 100 per cent inclusion probabilities, while the income polynomials range around 80 per cent. We find that less variation in precipitation, increased temperature and greater executive constraints reduce SO₂ concentrations. The only economic variable that registers as significant in the reduced-form station-level results is trade intensity. Here the evidence is decisive that trade reduces pollution.

Table 2. Reduced-form BMA results (by city)

	Dependent variable: log median SO ₂ concentrations, 5-year averages		
	$P \neq 0$	Posterior mean (S.D.)	Best model mean (S.D.)
Intercept	100.0	15.7627 (27.7715)	-4.5770*** (1.1233)
Trade	96.3	-2.1622 (0.8079)	-2.5691*** (0.7329)
Executive constraints	92.3	-0.1619 (0.0696)	-0.1916*** (0.0601)
Precipitation variation	91.6	42.1637 (18.6573)	51.0975*** (15.1028)
(GDP _{<i>t-3</i>}) ²	80.6	-2.7724 (2.1255)	-3.8403*** (1.2763)
(GDP _{<i>t-3</i>}) ³	80.2	0.5749 (0.4201)	0.7824*** (0.2412)
GDP _{<i>t-3</i>}	64.8	3.3708 (3.2758)	4.9447** (2.2508)
Year	40.7	-0.0102 (0.0143)	.
Education _{<i>t-3</i>}	20.2	0.0319 (0.0768)	.
Investment	18.5	-0.0035 (0.0089)	.
(GDP) ²	7.2	0.0197 (0.9318)	.
GDP	6.8	-0.0349 (1.3678)	.
(GDP) ³	6.8	-0.0028 (0.1857)	.
Helsinki	6.8	0.0121 (0.0723)	.
Capital intensity (H*K/L)	6.2	0.0044 (0.0338)	.
(Capital intensity, H*K/L) ²	4.8	-0.0001 (0.0013)	.
Temperature	3.8	-0.0013 (0.0169)	.
National population density	3.3	0.0027 (0.1189)	.
Observations			263
Adjusted R ²			0.303

Notes: $P \neq 0$ is the posterior inclusion probability that a regressor's posterior mean is different from zero. *, **, ***, indicate 90, 95, 99 per cent confidence levels.

The best single regression model selected by BMA at the station level has an adjusted R^2 of 0.240, and contains seven variables that exhibit at least weak evidence in terms of inclusion probabilities. The city-level results in table 2 are nearly identical to those at the station level, except that the previously weakly significant temperature variable is no longer relevant. Although the best model at the city level is based on fewer regressors and observations (both would lead one to expect a worse fit), its adjusted R^2 increases to 0.303. The improvement in explanatory power may result from the fact that the aggregated dataset is less prone to oversampling.

A major change in the results occurs when we aggregate the data to the country level. Table 3 no longer provides evidence that income has an influence on pollution. Nevertheless, all other variables that have been shown to be robustly related to pollution remain strongly significant and their posterior means are surprisingly stable. Executive constraints, trade, and local weather variations are central to explaining the country-level pollution variability. Interestingly, at the country level, education and technology (proxied by the year variable) now have high inclusion

Table 3. Reduced-form BMA results (by country)

	Dependent variable: log median SO ₂ concentrations, 5-year averages		
	$P \neq 0$	Posterior mean (S.D.)	Best model mean (S.D.)
Intercept	100.0	72.0870 (34.8970)	81.3495*** (28.5079)
Executive constraints	100.0	-0.2138 (0.0501)	-0.2193*** (0.0587)
Precipitation variation	99.9	59.9378 (14.7215)	60.4566*** (17.2538)
Temperature	98.9	-0.3451 (0.0987)	-0.3764*** (0.0993)
Education _{<i>t-3</i>}	97.1	0.4698 (0.1621)	0.5278*** (0.1500)
Trade	96.7	-1.6762 (0.6217)	-1.6468** (0.6317)
Year	90.0	-0.0369 (0.0179)	-0.0415*** (0.0147)
Helsinki	43.7	-0.2049 (0.2793)	-0.4946* (0.2674)
(GDP _{<i>t-3</i>}) ²	24.1	-0.1355 (0.5184)	.
(GDP _{<i>t-3</i>}) ³	15.1	0.0094 (0.1133)	.
GDP	11.7	0.0997 (0.5368)	.
GDP _{<i>t-3</i>}	10.8	0.0024 (0.6373)	.
(GDP) ²	8.6	0.0064 (0.2205)	.
(GDP) ³	8.5	0.0015 (0.0557)	.
National population density	7.3	0.0316 (0.2027)	.
Investment	6.3	-0.0005 (0.0037)	.
(Capital intensity, H*K/L) ²	5.0	-0.0001 (0.0009)	.
Capital intensity (H*K/L)	4.7	0.0002 (0.0157)	.
Observations			109
Adjusted R ²			0.465

Notes: $P \neq 0$ is the posterior inclusion probability that a regressor's posterior mean is different from zero. *, **, ***, indicate 90, 95, 99 per cent confidence levels.

probabilities, providing strong evidence that these candidate regressors belong in the true model. As we aggregate from the station to the city and finally to the country level, the adjusted R^2 of the best model systematically increases (although the number of observations drops from 623 to 109). While the adjusted R^2 is only 0.240 for the best model with station-level data, it nearly doubles to 0.465 at the country level.

The results for the regressors motivated by the theory-based approach of Antweiler *et al.* (2001) are presented in table 4. Antweiler *et al.* (2001) use station-level results as a benchmark, since their specification is the most extensive, theory-based empirical implementation of the EKC hypothesis. Alternative levels of aggregation (and hence different degrees of oversampling) generate remarkably stable outcomes in terms of posterior inclusion probabilities and posterior means. BMA identifies between six (station- and city-level) and eight (country-level) candidate regressors with weak to decisive evidence of a non-zero impact on pollution. Common across all levels of aggregation is that there exists no evidence of an EKC. Instead, the prominence of site-specific and political economy

variables carries over from our structural results at all levels of aggregation. There is strong and decisive evidence that non-economic factors, such as temperature, precipitation variation, and executive constraints, affect SO₂ concentrations in the same fashion as in the reduced-form BMA tables 1–3.

Of all the variables that receive positive evidence in table 4 at the station level, only executive constraints does not appear in the theory-specified Antweiler *et al.* (2001) model (they include the ‘Communist’ variable to capture political economy effects). Executive constraints thus remains highly significant not only in the reduced-form, but also in the structural analysis. This provides strong support for the Jones and Manuelli (2001) approach to pollution that emphasizes the political process, not income, as the driving force in the development/pollution relationship.

BMA produces a number of surprises. The major difference between the reduced-form specifications and the Antweiler theory-based results is that the trade intensity effect is lost entirely. While it does register as significant in Antweiler *et al.* (2001), BMA provides no evidence that trade intensity alone carries any explanatory power at the station, city, or country level. Nevertheless, the BMA results do suggest that trade plays an important indirect role in determining pollution since it is revealed that trade moderates the composition effect. Antweiler *et al.* (2001) find that trade dampens the pure EKC effect as the trade/income interactions in their regression are highly significant. In the BMA approach, in contrast, this income effect is found only at the station and city levels, and only in non-linear form. In the less unbalanced country-level dataset, BMA indicates that trade’s main role is to moderate the composition effect. The interaction between trade and capital intensity shows that the composition effect has a different impact on countries depending on their level of development. The greater the level of development – as proxied by the human-capital-augmented capital–labor ratio – the lower the implied concentrations for open economies.

The second trade-related variable that receives support in BMA at all levels of aggregation is the interaction between trade, income, and capital intensity. The positive estimate throughout provides strong evidence that more human/physical-capital-intensive countries have higher sulphur dioxide concentrations, even after we control for trade and income effects. This is because the three-way interaction between trade, income, and the human-capital-adjusted capital–labor ratio has a positive posterior mean. The relatively large role of the composition effect and the trade-based interactions suggests that countries do not follow a deterministic income–pollution path.

BMA does not uncover strong evidence for a pure composition effect, since capital intensity alone cannot be shown to affect pollution (in contrast to Antweiler *et al.*, 2001). In addition, since city-GDP/km² is not significant at any level of aggregation, BMA provides no evidence for a scale effect (the scale effect is only mildly significant in Antweiler *et al.*’s 2001 work). Oversampling does influence the strength of the technique effect (proxied by year), as BMA provides evidence that a technique effect reduces pollution at the country level. A similar pattern is observed in our reduced-form analysis, where the same variable gains explanatory power only at higher

Table 4. Structural BMA results

	Station				City			Country		
	<i>Antweiler et al. (2001) mean</i>	<i>P ≠ 0</i>	<i>Posterior mean (S.D.)</i>	<i>Best model mean (S.D.)</i>	<i>P ≠ 0</i>	<i>Posterior mean (S.D.)</i>	<i>Best model mean (S.D.)</i>	<i>P ≠ 0</i>	<i>Posterior mean (S.D.)</i>	<i>Best model mean (S.D.)</i>
Intercept	-4.299***	100.0	0.8530 (4.0590)	-0.0796 (1.0338)	100.0	4.1462 (16.4172)	-3.7002*** (0.4696)	100.0	65.7689 (35.1449)	77.6261** (28.4570)
Executive constraints	.	100.0	-0.1931 (0.0292)	-0.2014*** (0.0363)	99.1	-0.2058 (0.0530)	-0.2243*** (0.0568)	100.0	-0.2183 (0.0514)	-0.2329*** (0.0577)
Temperature	-0.056*	99.9	-0.2389 (0.0544)	-0.2220*** (0.0671)	28.7	-0.0418 (0.0758)	.	99.7	-0.3022 (0.0824)	-0.2897** (0.0920)
Precipitation variation	10.716*	99.9	33.3686 (7.3798)	33.0099*** (8.7721)	91.9	35.3218 (15.2463)	43.2826*** (12.6177)	99.5	53.0692 (14.6064)	54.9500** (16.3372)
Education _{t-3}	.	1.8	0.0007 (0.0094)	.	3.3	0.0030 (0.0225)	.	95.0	0.3920 (0.1515)	0.4289** (0.1437)
Year	.	3.9	-0.0003 (0.0020)	.	20.2	-0.0036 (0.0084)	.	86.4	-0.0338 (0.0181)	-0.0401** (0.0146)
Trade * relative GNP * relative H*K/L	0.924**	99.5	1.4380 (0.2966)	1.5591*** (0.3001)	98.6	1.7338 (0.4914)	1.9529*** (0.3989)	76.9	0.9260 (0.6205)	1.3144** (0.4051)
Trade * relative H*K/L	-2.121	99.5	-2.8599 (0.6483)	-3.1117*** (0.6106)	96.4	-3.2700 (1.0273)	-3.6332*** (0.8529)	70.0	-2.0269 (1.5041)	-3.0207*** (0.8725)
H*K/L * GNP _{t-3}	-0.386***	19.1	-0.0071 (0.0179)	.	28.0	-0.0246 (0.0457)	.	67.4	-0.0571 (0.0512)	-0.0890** (0.0298)
Trade	-3.216**	2.1	-0.0173 (0.1974)	.	10.4	-0.1607 (0.5524)	.	39.4	-0.7656 (1.1294)	.
(City-GDP/km ²) ² /1,000	-0.340	4.9	0.0099 (0.0568)	.	8.2	0.0305 (0.1235)	.	38.7	0.3227 (0.4869)	0.8462 (0.4316)

Table 4. *Continued*

Helsinki	0.016	3.0	-0.0064 (0.0468)	.	3.2	0.0102 (0.0752)	.	12.0	-0.0473 (0.1762)	.
City-GDP/km ²	-0.089*	4.9	0.0009 (0.0054)	.	15.9	0.0065 (0.0175)	.	11.7	0.0039 (0.0204)	.
Trade * (relative GNP) ²	-0.584**	76.0	-0.2870 (0.1718)	-0.3549*** (0.0651)	55.1	-0.1516 (0.1538)	-0.2904** (0.0896)	9.2	-0.0082 (0.0644)	.
Trade * relative GNP	2.614***	25.0	-0.2496 (0.4578)	.	19.8	-0.1432 (0.3356)	.	8.3	0.0106 (0.2195)	.
(GNP _{t-3}) ²	0.578***	3.6	-0.0012 (0.0214)	.	5.4	0.0164 (0.0908)	.	7.7	-0.0101 (0.0832)	.
(Capital intensity, H*K/L) ²	0.008	34.5	-0.0020 (0.0033)	-0.0049* (0.0025)	67.5	-0.0071 (0.0056)	-0.0107*** (0.0031)	7.4	-0.0003 (0.0018)	.
Trade * (relative H*K/L) ²	-0.176	2.3	-0.0032 (0.0280)	.	2.1	-0.0021 (0.0317)	.	7.4	-0.0132 (0.0753)	.
Capital intensity (H*K/L)	0.437*	7.6	0.0001 (0.0330)	.	6.2	-0.0044 (0.0446)	.	7.0	0.0044 (0.0395)	.
GNP _{t-3}	-0.228	4.2	0.0022 (0.1197)	.	6.1	-0.0682 (0.3705)	.	6.6	-0.0359 (0.2764)	.
Investment	.	35.0	-0.0067 (0.0102)	.	2.4	-0.0002 (0.0021)	.	3.7	-0.0003 (0.0029)	.
National population density	.	1.4	0.0025 (0.0421)	.	1.1	0.0023 (0.0695)	.	2.8	0.0107 (0.1198)	.
Communist * (GNP _{t-3}) ²	-8.806**	4.4	-0.0608 (0.9067)	.	1.5	0.0195 (0.3671)	.	2.4	-0.0554 (0.8900)	.
Communist * GNP _{t-3}	9.639**	42.6	1.7219 (2.3636)	.	3.8	0.0910 (0.6568)	.	1.9	0.0352 (0.9564)	.
Observations	2,555			653			273		115	
Number of effective regressors	18			7			6		9	
Adjusted R ²	0.144			0.235			0.301		0.479	

Notes: $P \neq 0$ is the posterior inclusion probability that a regressor's posterior mean is different from zero. *, **, ***, indicate 95, 99, 99.9 per cent confidence levels. Antweiler *et al.* (2001) do not report standard errors.

levels of aggregation. These findings are in line with Stern (2002) who finds evidence for the important role of negative time effects in explaining declining SO₂ concentrations.

Perhaps the most important result is that the best model chosen by BMA contains less than half of the 23 candidate regressors that have been motivated by the literature. At the station level, seven significant regressors account for about one-and-a-half times more variation in the dependent variable (adjusted $R^2 = 0.235$) than the 18 regressors (12 significant) suggested by the Antweiler *et al.* (2001) specification (adjusted $R^2 = 0.144$). This suggests that a number of regressors identified by Antweiler *et al.* (2001) may be significant only because the empirical strategy did not account for model uncertainty. The BMA estimates at different levels of aggregation are surprisingly stable, however, and their adjusted R^2 increases steadily from the station to the city to the country level, from 0.235 to 0.301 to 0.479, respectively (although the sample size declines). Also, the relevant regressors at the city and station levels are just about identical, although the country-level results do feature two additional regressors to explain pollution (year and education). The coefficient on education is counterintuitive just as in the reduced-form BMA results. It is supposed to proxy for the hypothesis that better-educated citizens demand better environmental quality. However, measures of education have been shown to be fragile in both growth regressions and in development accounting (see Krueger and Lindahl, 2001). Perhaps the same issues contaminate the effect of the regressors here.

6. Conclusion

This paper reexamines the evidence for an environmental Kuznets curve using the updated GEMS/AIRS data on SO₂ concentrations. The literature on the income–pollution relationship is characterized by unusual model uncertainty as both the number of proposed theories and the range of possible candidate regressors is large. We apply a theoretically founded method to address model uncertainty. Bayesian model averaging examines all models, weighs them by their relative quality, and then generates the probability that a candidate regressor is related to the dependent variable.

Our results are presented at three levels of aggregation. The station-level results are subject to severe oversampling as pollution from thousands of observations from local stations are linked to one and the same measure of income in a country. Hence we also aggregate the data to the city and country level. The results are remarkably robust. Political economy and site-specific variables explain a large share of the observed pollution. International trade is also shown to be robustly related to pollution. In our reduced-form analysis, trade is found to lower pollution. When the model is specified using full-fledged theories (Antweiler *et al.*, 2001), we show that trade has no direct effect, but that it moderates the composition effect. We provide evidence that as countries become richer and increase their physical and human capital, trade leads to cleaner environments. It unfortunately also implies that poor, labor-intensive, open economies experience increasing pollution levels.

Overall, we find only weak evidence for an EKC, which disappears when we address oversampling of the data or move to a fully specified theory-based approach. There may be several reasons the EKC fails to hold up in our work. The foremost, perhaps, is that many countries in the GEMS/AIRS data may already be on the flat or downward-sloping portion of the EKC during the sample period. Smulders *et al.* (2005) label these portions of the EKC the 'alarm phase' and the 'cleaning-up phase' that indicate a government response to public concerns. Given that the reduction in sulphur dioxide concentrations may also be based on governments reacting to their citizens' demands, it is not surprising that we find that policy variables such as executive constraints play a crucial role in determining pollution levels.

References

- Acemoglu, D., S. Johnson, and J.A. Robinson (2001), 'The colonial origins of comparative development: an empirical investigation', *American Economic Review* **91**: 1369–1401.
- Aghion, P. and P. Howitt (1998), 'Capital accumulation and innovation as complementary factors in long-run growth', *Journal of Economic Growth* **3**: 111–130.
- Anand, S. and S.M.R. Kanbur (1993), 'Inequality and development: a critique', *Development Economics* **41**: 19–43.
- Ang, B.W. (1999), 'Decomposition methodology in energy demand and environmental analysis', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham: Edward Elgar Publishers.
- Antweiler, W., B.R. Copeland, and M.S. Taylor (2001), 'Is free trade good for the environment?', *American Economic Review* **91**: 877–908.
- Arrow, K., B. Bolin, R. Costanza, P. Dasgupta, C. Folke, C.S. Holling, B.-O. Jansson, S. Levin, K.-G. Måler, C. Perrings, and D. Pimentel (1995), 'Economic growth, carrying capacity, and the environment', *Science* **268**: 520–521.
- Barrett, S. and K. Graddy (2000), 'Freedom, growth, and the environment', *Environment and Development Economics* **5**: 433–456.
- Barro, R.J. (1990), 'Government spending in a simple model of endogenous growth', *Journal of Political Economy* **98**: S103–S125.
- Barro, R.J. and J.-W. Lee (2000), 'International data on educational attainment: updates and implications', CID Working Paper No. 42.
- Begun, J. and T.S. Eicher (2007), 'In search of a sulphur dioxide environmental Kuznets curve: a Bayesian model averaging approach', University of Washington Working Paper UWEC 2007–18.
- Bovenberg, A.L. and S. Smulders (1995), 'Environmental quality and pollution-saving technological change in a two-sector endogenous growth model', *Journal of Public Economics* **57**: 369–391.
- Brock, W.A. and M.S. Taylor (2005), 'Economic growth and the environment: a review of theory and empirics', in P. Aghion and S.N. Durlauf (eds), *Handbook of Economic Growth*, Volume 1B, Amsterdam: Elsevier.
- Burnham, K.P. and D.R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*, 2nd edn, New York: Springer-Verlag.
- Constantini, V. and C. Martini (2007), 'A modified environmental Kuznetz curve for sustainable development assessment using panel data', FEEM Working Paper 148.
- Copeland, B.R. and M.S. Taylor (2003), *Trade and the Environment: Theory and Evidence*, Princeton, NJ: Princeton University Press.
- Cropper, M. and C. Griffith (1994), 'The interaction of population growth and environmental quality', *American Economic Review* **84**: 250–254.

- Dasgupta, S., B. Laplante, H. Wang, and D. Wheeler (2002), 'Confronting the environmental Kuznets curve', *Journal of Economic Perspectives* **16**: 147–168.
- De Groot, H.L.F., Z. Minliang, and C.A. Withagen (2004), 'The dynamics of China's regional development and pollution: an empirical investigation into the existence of an environmental Kuznets curve for China', *Environment and Development Economics* **9**: 507–537.
- De la Fuente, A. and R. Doménech (2001), 'Schooling data, technological diffusion, and the neoclassical model', *American Economic Review* **91**: 323–327.
- De la Fuente, A. and R. Doménech (2006), 'Human capital in growth regressions: how much difference does data quality make?', *Journal of the European Economic Association* **4**: 1–36.
- Deacon, R.T. and C.S. Norman (2006), 'Does the environmental Kuznets curve describe how individual countries behave?', *Land Economics* **82**: 291–315.
- Dijkgraaf, E. and H. Vollebergh (2005), 'A test for parameter homogeneity in CO₂ panel EKC estimations', *Environment and Resource Economics* **32**: 229–239.
- Draper, D. (1995), 'Assessment and propagation of model uncertainty (with discussion)', *Journal of the Royal Statistical Society, Series B* **57**: 45–97.
- Eicher, T.S., C. Papageorgiou, and A.E. Raftery (2007a), 'Bayesian model averaging in economics', Working Paper.
- Eicher, T.S., C. Papageorgiou, and O. Roehn (2007b), 'Unraveling the fortunes of the fortunate: an iterative Bayesian model averaging (IBMA) approach', *Journal of Macroeconomics* **29**: 494–514.
- Farnsworth, M.L., J.A. Hoeting, N.T. Hobbs, and M.W. Miller (2006), 'Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach', *Ecological Applications* **16**: 1026–1036.
- Fernández, C., E. Ley, and M.F.J. Steel (2001), 'Model uncertainty in cross-country growth regressions', *Journal of Applied Econometrics* **16**: 563–576.
- Fernández, C., E. Ley, and M.F.J. Steel (2002), 'Bayesian modeling of catch in a north-west Atlantic fishery', *Journal of the Royal Statistical Society: Series C* **51**: 257–280.
- Goklany, I. (2002), 'Comparing twentieth century trends in US and global agricultural water and land use', *Water International* **27**: 321–329.
- Grossman, G.M. and A.B. Krueger (1993), 'Environmental impacts of a North American free trade agreement', in P.M. Garber (ed.), *The US–Mexico Free Trade Agreement*, Cambridge, MA: MIT Press, pp. 13–56.
- Grossman, G.M. and A.B. Krueger (1995), 'Economic growth and the environment', *Quarterly Journal of Economics* **110**: 353–377.
- Harbaugh, W.T., A. Levinson, and D.M. Wilson (2002), 'Reexamining the empirical evidence for an environmental Kuznets curve', *Review of Economics and Statistics* **83**: 541–551.
- Hendry, D.F. and H.-M. Krolzig (2001), *Automatic Econometric Model Selection Using PcGets*, London: Timberlake Consultants Press.
- Heston, A., R. Summers, and B. Aten (2002), Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), http://pwt.econ.upenn.edu/php_site/pwt_index.php.
- Hjort, N.L. and G. Claeskens (2003), 'Frequentist model average estimators', *Journal of the American Statistical Association* **98**: 879–899.
- Hoekstra, R. and J.C.J.M. Van den Bergh (2002), 'Structural decomposition analysis of physical flows in the economy', *Environmental and Resource Economics* **23**: 357–378.
- Hoeting, J.A. (1994), 'Accounting for model uncertainty in linear regression', University of Washington Ph.D. dissertation.
- Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky (1999), 'Bayesian model averaging: a tutorial', *Statistical Science* **14**: 382–417.

- Hsiao, C. (1986), *Analysis of Panel Data*, Econometric Society Monographs No. 11, Cambridge: Cambridge University Press.
- Jaggers, K. and T.R. Gurr (1995), 'Polity III: regime change and political authority, 1800–1994', <http://privatewww.essex.ac.uk/~ksg/Polity.html>.
- Jeffreys, H. (1961), *Theory of Probability*, 3rd edn, Oxford: The Clarendon Press.
- Jones, L.E. and R.E. Manuelli (2001), 'Endogenous policy choice: the case of pollution and growth', *Review of Economic Dynamics* 4: 369–405.
- Kass, R.E. and A.E. Raftery (1995), 'Bayes factors', *Journal of the American Statistical Association* 90: 773–795.
- Kass, R.E. and L. Wasserman (1995), 'A reference Bayesian test for nested hypotheses and its relationship to the Schwartz criterion', *Journal of the American Statistical Association* 90: 928–934.
- Knack, S. and P. Keefer (1995), 'Institutions and economic performance: cross-country tests using alternative institutional measures', *Economics and Politics* 7: 207–227.
- Koop, G. and L. Tole (2006), 'An investigation of thresholds in air pollution–mortality effects', *Environmental Modelling and Software* 21: 1662–1673.
- Krueger, A.B. and M. Lindahl (2001), 'Education for growth: why and for whom?', *Journal of Economic Literature* 39: 1101–1136.
- Kuznets, S. (1955), 'Economic growth and income inequality', *American Economic Review* 45: 1–28.
- Leamer, E. (1978), *Specification Searches*, New York: Wiley.
- Leamer, E. (1983), 'Let's take the con out of econometrics', *American Economic Review* 73: 31–43.
- Leitão, A. (2006), 'Corruption and the environmental Kuznets curve: empirical evidence for sulfur', Working Paper, Universidade Católica Portuguesa.
- List, J.A. and C.A. Gallet (1999), 'The environmental Kuznets curve: does one size fit all?', *Ecological Economics* 31: 409–424.
- Madigan, D. and A.E. Raftery (1994), 'Model selection and accounting for model uncertainty in graphical models using Occam's window', *Journal of the American Statistical Association* 89: 1535–1546.
- Marshall, M.G. and K. Jaggers (2003), 'Polity IV Project: Political Regime Characteristics and Transitions, 1800–2003', <http://www.cidcm.umd.edu/inscr/polity/>.
- Martinez-Zarzoso, I. and A. Bengochea-Morancho (2004), 'Pooled mean group estimation for an environmental Kuznets curve for CO₂', *Economics Letters* 82: 121–126.
- Miller, A.J. (1984), 'Selection of subsets of regression variables (with discussion)', *Journal of the Royal Statistical Society Series A* 147: 389–425.
- Miller, A.J. (1990), *Subset Selection in Regression*, New York: Chapman & Hall.
- Millimet, D.L., J.A. List, and T. Stengos (2003), 'The environmental Kuznets curve: real progress or misspecified models?', *Review of Economics and Statistics* 85: 1038–1047.
- Mundlak, Y. (1978), 'On the pooling of time series and cross section data', *Econometrica* 46: 69–85.
- Panayotou, T. (1993), 'Empirical tests and policy analysis of environmental degradation at different stages of economic development', Working Paper WP238, Technology and Employment Programme, International Labor Office, Geneva.
- Panayotou, T. (1995), 'Environmental degradation at different stages of economic development', in I. Ahmed and J.A. Doeleman (eds), *Beyond Rio: The Environmental Crisis and Sustainable Livelihoods in the Third World*, London: Macmillan, pp. 13–36.
- Panayotou, T. (1997), 'Demystifying the environmental Kuznets curve: turning a black box into a policy tool', *Environment and Development Economics* 2: 465–484.

- Perman, R. and D.I. Stern (2003), 'Evidence from panel unit root and cointegration tests that the environmental Kuznets curve does not exist', *Australian Journal of Agricultural and Resource Economics* **47**: 325–347.
- Raftery, A.E. (1993), 'Bayesian model selection in structural equation models', in K.A. Bollen and J.S. Long (eds), *Testing Structural Equation Models*, Beverly Hills: Sage, pp. 163–180.
- Raftery, A.E. (1995), 'Bayesian model selection in social research (with discussion)', *Sociological Methodology* **25**: 111–196.
- Raftery, A.E. (1996), 'Approximate Bayes factors and accounting for model uncertainty in generalized linear models', *Biometrika* **83**: 251–266.
- Raftery, A.E. and Y. Zheng (2003), 'Long-run performance of Bayesian model averaging', *Journal of the American Statistical Association* **98**: 931–938.
- Roberts, J. and P. Grimes (1997), 'Carbon intensity and economic development 1962–91: a brief exploration of the environmental Kuznets curve', *World Development* **25**: 191–198.
- Rock, M.T. (1998), 'Freshwater use, freshwater scarcity, and socioeconomic development', *Journal of Environment and Development* **7**: 278–301.
- Sala-i-Martin, X.X. (1997), 'I just ran two million regressions', *American Economic Review* **87**: 178–183.
- Sala-i-Martin, X.X., G. Doppelhoffer, and R. Miller (2004), 'Determinants of long-term growth: a Bayesian averages of classical estimates (BACE) approach', *American Economic Review* **94**: 813–835.
- Selden, T.M. and D. Song (1994), 'Environmental quality and development: is there a Kuznets curve for air pollution?', *Journal of Environmental Economics and Environmental Management* **27**: 147–162.
- Shafik, N. and S. Bandyopadhyay (1992), 'Economic growth and environmental quality', World Bank Policy Research Working Paper WPS 904.
- Smulders, S., L. Bretschger, and H. Egli (2005), 'Economic growth and the diffusion of clean technologies: explaining environmental Kuznets curves', Economics Working Paper Series, Working Paper 05/42, WIF – Institute of Economic Research.
- Stern, D.I. (2002), 'Explaining changes in global sulfur emissions: an econometric decomposition approach', *Ecological Economics* **42**: 201–220.
- Stern, D.I. (2004), 'The rise and fall of the environmental Kuznets curve', *World Development* **32**: 1419–1439.
- Stern, D.I. (2005), 'Beyond the environmental Kuznets curve: diffusion of sulfur emissions abating technology', *Journal of Environment and Development* **14**: 101–124.
- Stern, D.I., M.S. Common, and E.B. Barbier (1996), 'Economic growth and environmental degradation: the environmental Kuznets curve and sustainable development', *World Development* **24**: 1151–1160.
- Stokey, N.L. (1998), 'Are there limits to growth?', *International Economic Review* **39**: 1–31.
- Torras, M. and J. Boyce (1998), 'Income, inequality, and pollution: a reassessment of the environmental Kuznets curve', *Ecological Economics* **25**: 147–160.

Appendix

Table A-1. Summary statistics

<i>Variable</i>	<i># of Observations</i>	<i>Mean</i>	<i>S.D.</i>	<i>Min.</i>	<i>Max.</i>
Log of median SO ₂	653	-4.883	1.052	-6.908	-2.181
GDP	623	1.328	0.848	0.108	2.673
GDP _{<i>t-3</i>}	623	1.279	0.819	0.104	2.626
City-GDP/km ²	653	6.666	8.061	0.103	57.565
GNP _{<i>t-3</i>}	653	1.220	0.789	0.113	2.476
Executive constraints	653	5.689	2.003	1.000	7.000
Investment	653	22.074	5.613	4.300	41.200
Trade	653	0.404	0.264	0.129	1.484
Capital intensity (H*K/L)	653	5.406	2.613	1.221	16.974
National population density	653	0.236	0.265	0.005	1.187
Education _{<i>t-3</i>}	653	7.758	3.041	1.222	11.806
Communist * GNP	653	0.038	0.126	0.000	0.688
Helsinki	653	0.033	0.159	0.000	1.000
Temperature	653	15.204	6.011	3.342	28.751
Precipitation variation	653	0.012	0.006	0.002	0.043

Note: The '*t-3*' subscript refers to an average of the past three years' data.

Table A-2. Correlations

	<i>Log of median SO₂</i>	<i>GDP</i>	<i>City- GDP/*km²</i>	<i>GNP_{t-3}</i>	<i>Executive constraints</i>	<i>Investment</i>	<i>Trade</i>	<i>Capital intensity (H*K/L)</i>	<i>National population density</i>	<i>Education_{t-3}</i>	<i>Communist* GNP</i>	<i>Helsinki</i>	<i>Temperature</i>
Log of median SO ₂	1.00												
GDP	-0.21	1.00											
City-GDP/km ²	0.22	0.46	1.00										
GNP _{t-3}	-0.20	0.99	0.44	1.00									
Executive Constraints	-0.28	0.67	0.27	0.65	1.00								
Investment	0.14	0.28	0.44	0.27	0.18	1.00							
Trade	-0.03	-0.06	-0.20	-0.07	0.09	0.15	1.00						
Capital intensity, (H*K/L)	0.04	0.55	0.28	0.57	0.38	0.45	0.23	1.00					
National population density	0.30	-0.24	0.28	-0.27	0.00	0.26	0.25	-0.03	1.00				
Education _{t-3}	-0.27	0.92	0.34	0.91	0.62	0.26	-0.03	0.36	-0.25	1.00			
Communist * GNP	0.16	-0.45	-0.20	-0.44	-0.48	0.04	-0.21	-0.46	0.08	-0.29	1.00		
Helsinki	-0.10	0.19	0.04	0.23	0.13	0.15	0.17	0.27	-0.06	0.18	-0.07	1.00	
Temperature	-0.16	-0.47	-0.17	-0.49	-0.30	-0.31	-0.23	-0.43	0.13	-0.49	-0.04	-0.25	1.00
Precipitation variation	-0.02	-0.57	-0.31	-0.59	-0.36	-0.42	-0.10	-0.52	0.18	-0.53	0.16	-0.22	0.44